

Review

Hagen Hofmann*

Single-molecule spectroscopy of unfolded proteins and chaperonin action

Abstract: In the past decade, single-molecule fluorescence techniques provided important insights into the structure and dynamics of proteins. In particular, our understanding of the heterogeneous conformational ensembles of unfolded and intrinsically disordered proteins (IDPs) improved substantially by a combination of FRET-based single-molecule techniques with concepts from polymer physics. A complete knowledge of the forces that act in unfolded polypeptide chains will not only be important to understand the initial steps of protein folding reactions, but it will also be crucial to rationalize the coupling between ligand-binding and folding of IDPs, and the interaction of denatured proteins with molecular chaperones in the crowded cellular environment. Here, I give a personalized review of some of the key findings from my own research that contributed to a more quantitative understanding of unfolded proteins and their interactions with molecular chaperones.

Keywords: IDP; polymer theory; protein collapse; protein folding; scaling laws; single-molecule FRET.

*Corresponding author: Hagen Hofmann, Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland, e-mail: h.hofmann@bioc.uzh.ch

Introduction

Starting with the classic biochemical approaches that had driven the field of protein folding and dynamics for decades, protein folding research has developed substantially in the past decade (Lindorff-Larsen et al., 2011; Chung et al., 2012; Dill and MacCallum, 2012; Schuler and Hofmann, 2013). One of the major goals of the present research is a detailed physical understanding of the self-organization of proteins through an intimate combination of theory, molecular simulations, and advanced experimental methods.

However, experiments and simulations have to close the gap to theory. Already in the late 1980s and early 1990s, theories of protein folding and protein collapse have been developed (Dill, 1985; Bryngelson and Wolynes, 1987, 1990; Shakhnovich and Gutin, 1989, 1990; Stigter and Dill, 1990; Alonso and Dill, 1991; Bryngelson et al., 1995; Dill and Stigter, 1995) based on the statistical physics of polymers and spin glasses that reach back to the original work of Flory (1949) and de Gennes (1975). To close this gap and quantitatively use these theories, experimental quantities that provide structural information directly, such as intramolecular distances, are needed. Single-molecule fluorescence spectroscopy provides this information with high precision and at a subpopulation specific level (Schuler, 2007; Schuler and Hofmann, 2013). Additionally, not only distances but the timescales of their interconversion are available from picoseconds up to hours and days (Schuler and Hofmann, 2013). Both types of information, intramolecular distances and the timescales of their interconversion were particularly important to understand the behavior of unfolded polypeptide chains (Hoffmann et al., 2007; Ziv and Haran, 2009; Müller-Späth et al., 2010; Hofmann et al., 2012; Soranno et al., 2012). Although a detailed understanding of unfolded proteins by the combination of experiment, theory, and simulation seems to be of rather theoretical interest, the realization that intrinsically disordered proteins (IDPs) are important regulatory hubs in intracellular processes, such as transcription activation and signaling (Ferreon et al., 2013), established a direct link between the physics of unfolded protein chains and the biological function of IDPs. Similarly, the function of molecular chaperones is difficult to rationalize without a quantitative understanding of the properties of their substrates, which often are denatured or even misfolded.

For example, it has been suggested that the confinement of proteins inside a cavity, such as the one formed by the heptameric rings of GroEL and GroES (Thirumalai and Lorimer, 2001; Hartl and Hayer-Hartl, 2002; Fenton and Horwich, 2003; Bukau et al., 2006; Hartl et al., 2011), can accelerate protein folding

reactions (Betancourt and Thirumalai, 1999; Klimov et al., 2002; Jewett et al., 2004; Hayer-Hartl and Minton, 2006; Tang et al., 2006). However, results from experiments and simulation show that this accelerating effect can be counteracted by interactions of the denatured substrate with the GroEL-GroES cavity wall (Hofmann et al., 2010). In the following, some of the basic principles of single-molecule fluorescence techniques are reviewed and examples from my own research are discussed to show how single-molecule experiments can be used to understand the polymer physical properties of unfolded and intrinsically disordered proteins.

Distances in unfolded proteins from single-molecule Förster resonance energy transfer

Förster resonance energy transfer (FRET) at the single-molecule level is ideally suited to obtain distances in unfolded proteins. Even in the presence of an excess of

folded proteins, the subpopulation of unfolded proteins still can be identified (Figure 1A). In the typical single-molecule fluorescence experiment, the protein labeled with a donor and an acceptor fluorophore diffuses randomly in and out of the confocal volume produced by a laser beam that passed an objective with high numerical aperture (Figure 1B).

Once a doubly labeled molecule stochastically enters the confocal volume, the donor fluorophore is excited by the laser light. To return to its ground state, the donor either emits a photon or it transfers its excitation energy to the acceptor. In its classical description, the latter process (FRET) originates from the resonance of the transition dipole moments of donor and acceptor fluorophores. The efficiency of the resulting energy transfer (E) from D to A depends on the inverse 6th power of the D–A distance (r) and the characteristic distance R_0 (Figure 2A), which can be calculated from the spectroscopic properties of D and A. The typical values of R_0 make single-molecule FRET suitable for obtaining distance information in the range

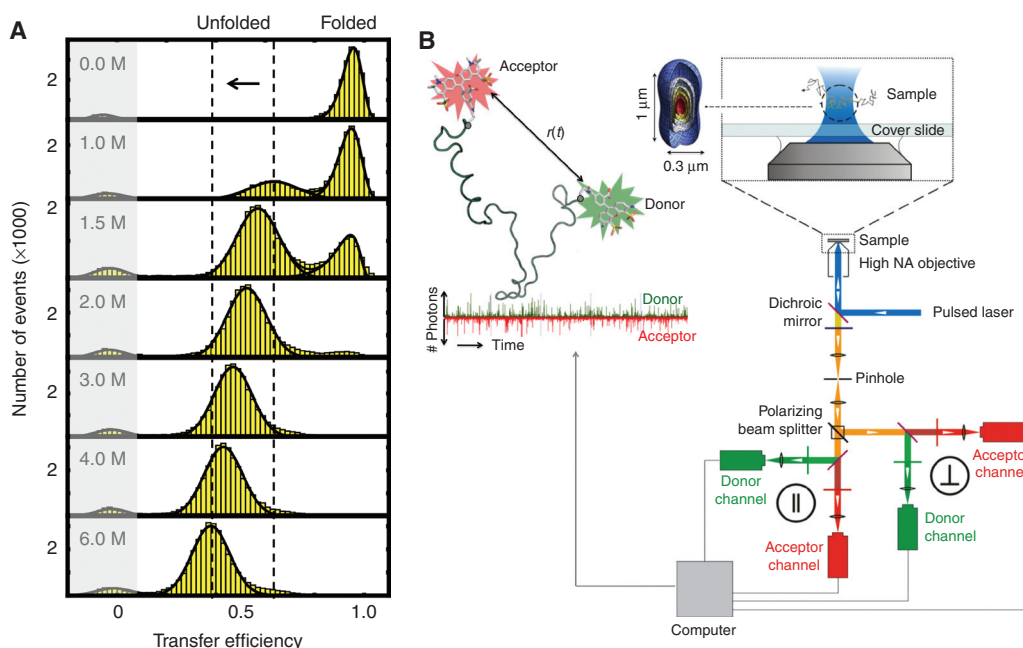


Figure 1 Collapse of unfolded CspTm (A) mapped with a confocal single-molecule fluorescence setup.

(A) Transfer efficiency histograms of doubly labeled CspTm at different concentrations of the denaturant guanidinium chloride (GdmCl) (Hofmann et al., 2012). Three peaks are observed. The peak at low transfer efficiencies corresponds to molecules without an active acceptor fluorophore. The peak at high transfer efficiencies corresponds to folded molecules, whereas the peak at intermediate transfer efficiency corresponds to unfolded CspTm molecules. With increasing concentrations of GdmCl, the transfer efficiency distribution of unfolded CspTm shifts continuously to lower transfer efficiencies (black arrow) suggesting an expansion of the unfolded chain with increasing denaturant concentrations. (B) Confocal single-molecule experiment on freely diffusing molecules. In this example of a four-channel instrument, the fluorescence signal is first separated by polarization and then by wavelength into the detection channels corresponding to emission from donor and acceptor chromophores. Data can either be taken on molecules freely diffusing in solution at very low concentration and observed while they diffuse through the confocal detection volume, or data on immobilized molecules can be recorded by sample or laser scanning.

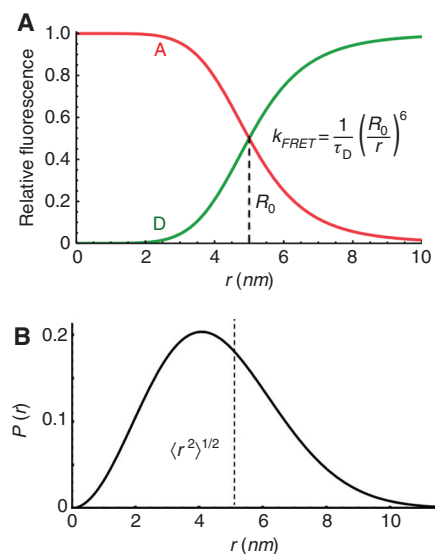


Figure 2 Distance-intensity relation for FRET and Gaussian chain model.

(A) Change of the fluorescence intensity, in other words the number of photons, for the donor (green) and acceptor (red) fluorophore as function of the distance between them. The Förster radius R_0 (dashed line) is the distance at which the transfer efficiency is 0.5. (B) Distance distribution function for a Gaussian chain calculated according to Equation (3). The root-mean-squared end-to-end distance $\langle r^2 \rangle^{1/2}$ is indicated (dashed line).

of ~2 nm to 10 nm. Importantly, the transfer efficiency (E) can be obtained directly from the measured number of acceptor and donor photons (n_A , n_D) for each single molecule via

$$E(r) = \frac{R_0^6}{R_0^6 + r^6} = \frac{n_A}{n_A + n_D}. \quad (1)$$

Before calculating transfer efficiencies from n_A and n_D , the detected photon counts in both detectors have to be corrected for their relative molecular brightness, background, detection efficiencies of the detectors, and the direct excitation of the acceptor (Schuler, 2007). Once the correct transfer efficiency is known, Equation (1) relates E to a single distance (r). However, in contrast to folded proteins, the distance between D and A in unfolded proteins and IDPs is not a fixed quantity. Instead, the D–A distance will fluctuate at a timescale of hundreds of nanoseconds because of the diffusive dynamics of the polypeptide chain (Bieri et al., 1999; Neuweiler et al., 2003; Möglich et al., 2006; Fierz et al., 2007; Nettels et al., 2007; Neuweiler et al., 2009; Soranno et al., 2012), which leads to a broad distribution of distances $P(r)$. A common misunderstanding of single-molecule FRET

experiments is that a broad distribution of distances also will lead to a broad distribution of transfer efficiencies. Instead, it is important to compare the timescale of the distance fluctuations with the time resolution of the experiment. The typical residence time of a protein molecule in the confocal spot is ~1 ms. During this time, several hundreds of photons can be recorded from the donor and acceptor dye, which finally leads to an average time between successive photons (interphoton time) of approximately 1–10 μ s. However, this timescale is much slower than the distance dynamics of unfolded proteins, and the polypeptide chain will sample a large part of its conformational space during the interphoton time. This fast conformational averaging leads to narrow transfer efficiency distributions whose mean value $\langle E \rangle$ is related to the distance distribution $P(r)$ via

$$\langle E \rangle = Z^{-1} \int_0^L E(r) P(r) dr \quad \text{with} \quad Z = \int_0^L P(r) dr. \quad (2)$$

Here, L is the contour length of the unfolded protein, which is given by the length of one peptide bond ($b=0.38$ nm) times the number of peptide bonds (N_{bonds}). In contrast, the width of the transfer efficiency distribution is dominated mainly by the uncertainty in the determined transfer efficiency, which results from the small number of detected photons (shot noise). To use Equation (2), several polymer models can be used to estimate $P(r)$. Certainly one of the simplest models is the Gaussian chain model whose distance distribution is given by

$$P(r) = 4\pi r^2 \left(\frac{3}{2\pi \langle r^2 \rangle} \right)^{3/2} \exp \left(-\frac{3r^2}{2\langle r^2 \rangle} \right) \quad (3)$$

where $\langle r^2 \rangle$ is the mean-squared distance between D and A (Figure 2B). It has been shown that other polymer models, such as the worm-like chain model (Thirumalai and Ha, 1998) or an excluded-volume chain model with interactions (Sanchez, 1979) yield similar results (Hofmann et al., 2012). The only free parameter in the Gaussian chain model is $\langle r^2 \rangle$, which can be obtained from $\langle E \rangle$ by numerically solving Equation (2). Within the framework of the simple polymer model described, the mean-squared distance between the attached fluorophores is related to the radius of gyration of the chain, which is a global measure of its dimension and an important parameter in all polymer theories. In the following section, examples of how the described single-molecule approach has been used in the past to understand generic properties of unfolded and intrinsically disordered proteins are discussed.

The effect of solvation on unfolded proteins and IDPs

Unfolded proteins compact upon transfer from high to low concentrations of denaturants, such as urea or guanidinium chloride (GdmCl) (Schuler et al., 2002; Sinha and Udgaonkar, 2005; Magg et al., 2006; Sherman and Haran, 2006; Merchant et al., 2007) (Figure 1A).

This compaction, often referred to as collapse of the polypeptide chain, has been identified in a multitude of single-molecule studies using Förster resonance energy transfer (smFRET) as a reporter of the dimension of unfolded proteins (Deniz et al., 2000; Schuler et al., 2002; Kuzmenkina et al., 2005; Sherman and Haran, 2006; Hoffmann et al., 2007; Merchant et al., 2007). Using fast fluorescence correlation techniques, the timescale of the conformational fluctuations in unfolded, and intrinsically disordered proteins has been found to be in the order of $\sim 0.1 \mu\text{s}$ (Nettels et al., 2007), which suggests the absence of a large free energy barrier between the expanded and collapsed unfolded state ensemble.

However, what are the molecular driving forces of this compaction? Already in the first experimental description of the collapse process using single-molecule FRET, it has been suggested that the binding of denaturant molecules to the polypeptide chain is responsible for the observed expansion in solutions containing increasing amounts of GdmCl (Schuler et al., 2002; Sherman and Haran, 2006). Until recently (Ziv and Haran, 2009; Hofmann et al., 2012), however, a direct link between the free energy of solvation of the polypeptide chain and its dimension, expressed by the radius of gyration, was missing. It was the combination of single-molecule FRET experiments with polymer theory that provided a quantitative explanation for the

denaturant-driven expansion of unfolded and intrinsically disordered proteins.

The first important step toward a quantitative understanding of the collapse process was taken by Ziv and Haran (2009), who analyzed a large data set of single-molecule FRET experiments in which the mean transfer efficiency of unfolded proteins has been obtained over a broad range of denaturant concentrations. The authors used a mean-field polymer theory (Sanchez, 1979) to convert the experimentally obtained mean transfer efficiency directly $\langle E \rangle$ into a mean interaction energy between the amino acids in the unfolded proteins. After computing free energies, they found that the change in the unfolded-state free energy upon addition of denaturant correlates with the change in the free energy of folding. Correspondingly, they concluded that a major part of the free energy change for folding originates from the free energy change of collapse. Although this thermodynamic link between folding and collapse has been used already in theory (Alonso and Dill, 1991), it was the first time that it was demonstrated experimentally.

Recently, Hofmann et al. (2012) extended the analysis of Ziv and Haran (2009) to a broad range of proteins with very different sequence composition, ranging from nearly charge-balanced sequences up to highly charged IDPs, such as prothymosin α (Hofmann et al., 2012). Despite largely different sequence properties of the investigated proteins, the authors found that the change in the mean interaction energy between the amino acids in these unfolded proteins was in quantitative agreement with the change in the free energy of solvation of the chain (Figure 3A), a quantity that can be calculated from the solvation free energies of the free amino acids (Nozaki and Tanford, 1970). This observation provided a simple

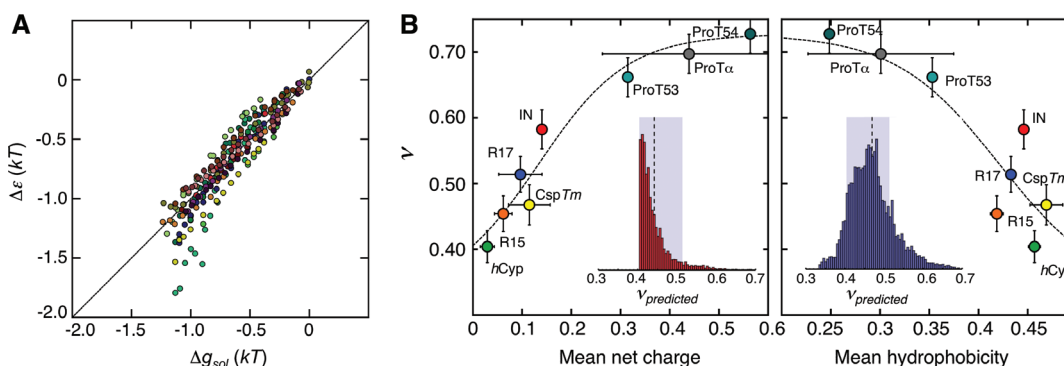


Figure 3 Interaction energies and scaling exponents for unfolded and intrinsically disordered proteins (Hofmann et al., 2012).

(A) Correlation of the difference in the interaction energy between water and GdmCl ($\Delta\epsilon$) for six proteins and its variants with the free energy of solvation (Δg_{sol}). (B) Scaling exponents obtained for six unfolded proteins, including two IDPs, as function of the mean net charge (left) and mean hydrophobicity (right) of the sequence. Insets: predicted scaling exponents of the unfolded state for 10 905 protein sequences from the PDB.

explanation for the correlation found by Ziv and Haran (2009). When adding denaturants, such as GdmCl, the release of the solvation free energy in the unfolded protein provides the driving force for the expansion of unfolded proteins and IDPs, and shifts the equilibrium between folded and unfolded proteins toward the side of unfolded proteins.

However, this simple model cannot explain the length-scaling of the radius of gyration of unfolded proteins and IDPs under physiological conditions. Scaling laws (Flory, 1949; Grosberg and Kuznetsov, 1992) relate the radius of gyration of a polymer (R_g) to its length, in other words, the number of peptide bonds (N), via $R_g \propto N^\nu$. Importantly, the exponents of these scaling laws provide information on the type of forces that dominate the size of polymers. For example, the size of a polymer without intrachain interactions will be determined by its excluded volume – by the volume of the amino acids in case of proteins. Such a chain will have a scaling exponent of 0.6. In contrast, if intrachain interactions dominate, the chain will collapse, and the scaling exponent reduces to 0.33. Between the two limits, under conditions where the repulsive and attractive contributions to the free energy of the chain are exactly balanced, the scaling exponent will reach a value of one half. These conditions are called Θ -conditions. Experiments using small-angle X-ray scattering (SAXS) and dynamic light scattering (DLS) showed that unfolded proteins in high concentrations of denaturant indeed show a scaling exponent close to 0.6 (Uversky, 2002; Kohn et al., 2004). Hence, under these conditions, the dominant interactions result from the excluded volume of the amino acids.

In contrast, the scaling exponents of unfolded proteins under physiological conditions are more difficult to obtain. Hofmann et al. (2012) found that the scaling exponents of their investigated proteins converge to a value close to 0.6 at high denaturant concentrations, in line with SAXS and DLS experiments (Uversky, 2002; Kohn et al., 2004). However, the scaling exponents strongly diverge from this value when the denaturant concentration is lowered. In the absence of GdmCl – in native conditions – the authors found that the scaling exponents depend sensitively on the sequence composition (Figure 3B). For example, the scaling exponents rise with increasing net charge of the sequences, but they lower with increasing hydrophobicity, suggesting that the balance of polymer-solvent and polymer-polymer interaction is altered by the chemical properties of the amino acid side chains. Most importantly, the average scaling exponent of the four moderately charged foldable protein sequences was 0.46 ± 0.05 under native

conditions, in other words, close to the ideal value of one half. This suggests that water is close to a Θ -solvent for these unfolded proteins.

What do these results imply for protein folding? Especially during the early stages of the folding process, many interactions have to be sampled to find the correct contacts that incrementally decrease the energy of the protein. Simulations based on simple models predict that unfolded chains close to the Θ -regime can accomplish this optimization process more efficiently than chains that are in the completely collapsed globule regime (Camacho and Thirumalai, 1993; Thirumalai et al., 2010). Hence, folding has been suggested to be most efficient close to the Θ -regime (Camacho and Thirumalai, 1993; Thirumalai et al., 2010) and the discussed results imply that natural protein sequences are indeed close to this regime.

The effect of charges on unfolded proteins and IDPs

Although substantial progress has been made in the past by applying polymer concepts to unfolded proteins and IDPs, the description of unfolded proteins by one mean-field interaction energy remains incomplete. Proteins usually have a complex sequence composition consisting of 20 different types of amino acids. Averaging over all possible types of interactions by extracting one mean interaction energy (Haran, 2012; Hofmann et al., 2012) is only a crucial first step for understanding the behavior of polypeptide chains. This aspect gained particular attention for IDPs, proteins that have an unusual high amount of charged amino acids (Uversky et al., 2000).

A first qualitative observation of the effect of charge interactions on the dimension of an unfolded protein was obtained by Hofmann et al. (2008). Using single-molecule FRET, the authors found a surprising discrepancy between the radius of gyration of the unfolded ribonuclease inhibitor barstar in low concentrations of urea and GdmCl. In the neutral denaturant urea, unfolded barstar was significantly more expanded than in the ionic denaturant GdmCl. Using a non-denaturing salt (KCl), the authors showed that unfolded barstar indeed collapsed with increasing ionic strength of the solution, which points toward the presence of substantial electrostatic repulsions in the chain at low ionic strength. Even though this explanation is physically intuitive, it raises an important question. Which sequence properties determine the strength and the sign of electrostatic interactions in unfolded proteins? Müller-Spätth et al. (2010) investigated this aspect by using single-molecule fluorescence spectroscopy to investigate four

different polypeptide sequences (three IDP sequences and one foldable sequence) with largely different net charge (Müller-Späth et al., 2010). For the highly charged IDP sequences of prothymosin α (N-terminal part: -14, C-terminal part: -27) and the N-terminal domain of HIV integrase (-4), the authors found a continuous compaction of the polypeptide chains with decreasing concentrations of the denaturant GdmCl (Figure 4C,D).

From the discussion in the preceding sections, this behavior is expected because of the change in the free energy of solvation of the polypeptide chain. However, at very low concentrations of GdmCl, the dimensions of the investigated IDPs expand again. Experiments using the neutral denaturant urea, instead of the salt GdmCl, did not reveal such an expansion at low concentrations of denaturant, suggesting that repulsive electrostatic interactions between the charged amino acids in the chain are responsible for the effect in GdmCl, similar to the effect found for the unfolded state of barstar (Hofmann et al.,

2008). Two types of polymer theories were able to describe quantitatively the observed effect, a polyelectrolyte theory (Ha and Thirumalai, 1992) in which the electrostatics of the chain only depend on the net charge of the sequence, and a polyampholyte theory (Higgs and Joanny, 1991), which takes repulsive and attractive interactions approximately into account. Importantly, the foldable protein sequence, a destabilized variant of the cold shock protein from *Thermotoga maritima* (CspTm) with a low net charge (-2), did not exhibit a reexpansion of its unfolded state at low GdmCl concentrations. In contrast, unfolded CspTm even showed an additional collapse at low ionic strength (Figure 4A), which is a clear indication that attractive electrostatic interactions start to dominate. Correspondingly, only polyampholyte theory, which explicitly takes repulsive and attractive electrostatic interactions into account, was able to explain this effect quantitatively. The counter-intuitive observation that the three IDP sequences expand with decreasing ionic strength, whereas the foldable protein sequence of CspTm collapses with decreasing ionic strength, could only be understood in terms of a theory that explicitly takes repulsive and attractive charge effects into account.

Interaction of an unfolded protein with GroEL-GroES

From the two preceding sections it is clear that the dimensions of unfolded proteins and IDPs depend sensitively on the solution conditions and on the amino acid composition of the chain. However, in the crowded environment of a living cell, cellular factors, such as chaperones and machineries that translocate and degrade proteins, can exert additional forces on polypeptide chains, which will have a large effect on the conformation and dynamics of polypeptide chains (Hartl et al., 2011). The complexity of these factors together with their promiscuity (Houry et al., 1999; Kerner et al., 2005; Stan et al., 2006) often complicate a quantitative investigation of their effects on unfolded proteins and protein folding. The use of external fluorophores and the ability to discriminate between subpopulations with different interdyne distances makes single-molecule fluorescence spectroscopy to a versatile tool for the investigation of the conformations and dynamics of proteins in more complex systems such as molecular chaperones.

Hofmann et al. (2010) used a single-molecule FRET approach to investigate the chaperonin GroEL-GroES. This remarkable molecular machine binds nonnative proteins and allows them to fold within a cavity formed

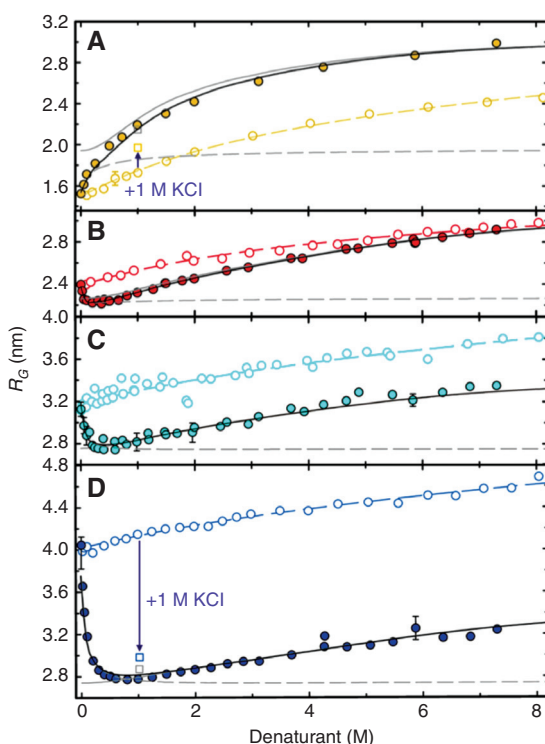


Figure 4 The effect of charges on the dimensions of unfolded proteins and IDPs (Müller-Späth et al., 2010).

Radii of gyration as function of the denaturant concentration (GdmCl: filled symbols, urea: open symbols) for unfolded CspTm (A), and three IDP-sequences: N-terminal domain of HIV-integrase (B), N-terminal domain of prothymosin α (C), and C-terminal domain of prothymosin α (D). The difference between urea and GdmCl implies the presence of electrostatic interactions. Black solid lines are fits with polyampholyte theory (Higgs and Joanny, 1991).

by the heptameric rings of GroEL and GroES. Based on simple polymer models and molecular simulations, it was expected that the confinement of proteins inside a cavity, such as the one formed by the heptameric rings of GroEL and GroES (Thirumalai and Lorimer, 2001; Hartl and Hayer-Hartl, 2002; Fenton and Horwich, 2003; Bukau et al., 2006; Hartl et al., 2011), can accelerate protein folding reactions by an entropic destabilization of the unfolded chain (Betancourt and Thirumalai, 1999; Klimov et al., 2002; Jewett et al., 2004; Hayer-Hartl and Minton, 2006; Tang et al., 2006). However, the cavity is only slightly larger than the folded structure of typical proteins known to interact with the chaperonin. The large volume of unconfined unfolded protein chains compared to the size of the cavity therefore raised the question of whether and how such strong confinement affects the folding reaction (Zhou and Dill, 2001; Klimov et al., 2002; Baumketner et al., 2003; Takagi et al., 2003; Jewett et al., 2004; Tang et al., 2006; Farr et al., 2007; Mittal and Best, 2008).

By placing the FRET-fluorophores at different positions in the two-domain substrate protein rhodanese, Hofmann et al. (2010) used single-molecule FRET to map the conformation of the N-terminal (N-variant) and C-terminal domain (C-variant), and the conformation of the linker that connects both domains (L-variant). Binding of unfolded rhodanese to GroEL led to very broad transfer efficiency distributions. However, a broad single-molecule transfer efficiency histogram does not originate necessarily from a broad distance distribution. Particularly for the interaction of rhodanese with GroEL, control experiments revealed high-fluorescence anisotropies, which suggests that the fluorophores attached to rhodanese are not free to rotate in complex with GroEL. Instead, a previous study of Hillger et al. (2008) showed that a distribution of different static fluorophore orientations was responsible for the observed broad transfer efficiency histograms, which complicates the accurate determination of distances. Although intrachain distances in unfolded rhodanese were difficult to extract when the substrate was bound to GroEL, the folding kinetics still could be obtained.

Important for an evaluation of the effect of confinement on the folding reaction of rhodanese is the direct comparison of the chaperonin-mediated folding with the autonomous reaction in the absence of GroEL-GroES. Especially for stringent chaperone substrates, this comparison is complicated by their high aggregation propensity. However, the low concentrations used in single-molecule experiments (50 pM) effectively prevent aggregation, and a direct comparison between autonomous and chaperonin-mediated folding was possible. Surprisingly,

Hofmann et al. (2010) found that the effect of GroEL-GroES on the folding rate constants of rhodanese was marginal. Only the folding of the C-terminal domain of rhodanese was affected.

However, in contrast to the expectation from simple polymer theories, folding was not accelerated but decelerated (Figure 5). As an origin for this deceleration, Hofmann et al. (2010) suggested the presence of attractive interactions between the unfolded substrate and the wall of GroEL-GroES. Indeed, in a recent study (Sirur and Best, 2013), the authors used coarse-grained molecular dynamics simulations to show that the interactions of rhodanese with the GroEL-GroES cavity wall can indeed explain the experimentally observed deceleration of rhodanese folding if strong interactions of the unfolded chain with the cavity wall were imposed. The results of Hofmann et al. (2010) and Sirur and Best (2013) show that simple polymer models also have limits.

Although these theories are very successful in the description of unfolded and intrinsically disordered proteins, they miss many of the specific aspects that might be important to describe the interaction of unfolded proteins with large molecular assemblies, such as the chaperonin GroEL-GroES. Nevertheless, these theories are extremely helpful in formulating predictions on the basis of simple physical effects that finally allow a more rigorous interpretation of experimental results.

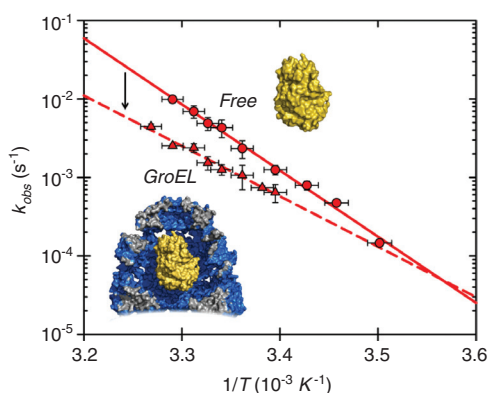


Figure 5 Direct comparison of folding dynamics outside and inside the GroEL/ES chaperonin cavity from single molecule fluorescence spectroscopy (Hofmann et al., 2010).

Compared to folding of the C-terminal domain of rhodanese free in solution (circles), the rate constants of folding inside the single-ring GroEL/ES cavity are reduced, suggesting that interactions between denatured rhodanese and the cavity walls or internal friction can lead to slowed folding inside the chaperonin cage. Insets: Schematic surface representations of native rhodanese (top) and native rhodanese inside the GroEL-GroES cavity.

Summary

In the past decade, the combination of single-molecule fluorescence techniques with polymer theoretical models has contributed significantly to our understanding of unfolded and intrinsically disordered proteins. However, the success of this combination was, at least in part, possible because of the relative simplicity of the investigated systems. More complex problems, such as the effect of molecular chaperones on protein folding reactions, are difficult to capture by generic polymer theories, because specific interactions might dominate the observed effects. Of course, the discrepancy between theory and experiment is very likely to increase with increasing complexity of the system under investigation, for example if the processes of folding and binding are investigated in the crowded environment of a living cell. Therefore, in the

future, the development of new theories with broader applicability needs to go hand in hand with molecular dynamics simulations and a further development of the experimental toolbox of single-molecule techniques to obtain a more accurate and physics-based understanding of the complex biochemical processes inside a living cell.

Acknowledgments: A part of the described research was initiated while I was a member of the GRK 1026 ‘Conformational transitions in macromolecular interactions,’ which I would like to thank in this particular way for the intellectual and financial support during the wonderful time we had together at the Martin-Luther University Halle-Wittenberg.

Received February 14, 2014; accepted March 4, 2014; previously published online March 11, 2014

References

- Alonso, D.O. and Dill, K.A. (1991). Solvent denaturation and stabilization of globular proteins. *Biochemistry* 30, 5974–5985.
- Baumketner, A., Jewett, A., and Shea, J.E. (2003). Effects of confinement in chaperonin assisted protein folding: rate enhancement by decreasing the roughness of the folding energy landscape. *J. Mol. Biol.* 332, 701–713.
- Betancourt, M.R. and Thirumalai, D. (1999). Exploring the kinetic requirements for enhancement of protein folding rates in the GroEL cavity. *J. Mol. Biol.* 287, 627–644.
- Bieri, O., Wirz, J., Hellrung, B., Schutkowski, M., Drewello, M., and Kiefhaber, T. (1999). The speed limit for protein folding measured by triplet-triplet energy transfer. *Proc. Natl. Acad. Sci. USA* 96, 9597–9601.
- Bryngelson, J. and Wolynes, P. (1987). Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. USA* 84, 7524–7528.
- Bryngelson, J. and Wolynes, P. (1990). A simple statistical field-theory of heteropolymer collapse with application to protein folding. *Biopolymers* 30, 177–188.
- Bryngelson, J., Onuchic, J., Socci, N., and Wolynes, P. (1995). Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* 21, 167–195.
- Bukau, B., Weissman, J., and Horwich, A. (2006). Molecular chaperones and protein quality control. *Cell* 125, 443–451.
- Camacho, C. and Thirumalai, D. (1993). Kinetics and thermodynamics of folding in model proteins. *Proc. Natl. Acad. Sci. USA* 90, 6369–6372.
- Chung, H.S., McHale, K., Louis, J.M., and Eaton, W.A. (2012). Single-molecule fluorescence experiments determine protein folding transition path times. *Science* 335, 981–984.
- de Gennes, P. (1975). Collapse of a polymer chain in poor solvents. *J. Phys. Lett.* 3, L55–L57.
- Deniz, A.A., Laurence, T.A., Beligere, G.S., Dahan, M., Martin, A.B., Chemla, D.S., Dawson, P.E., Schultz, P.G., and Weiss, S. (2000). Single-molecule protein folding: diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proc. Natl. Acad. Sci. USA* 97, 5179–5184.
- Dill, K.A. (1985). Theory for the folding and stability of globular proteins. *Biochemistry* 24, 1501–1509.
- Dill, K. and Stigter, D. (1995). Modeling protein stability as heteropolymer collapse. *Adv. Protein. Chem.* 46, 59–104.
- Dill, K.A. and MacCallum, J.L. (2012). The protein-folding problem, 50 years on. *Science* 338, 1042–1046.
- Farr, G.W., Fenton, W.A., and Horwich, A.L. (2007). Perturbed ATPase activity and not ‘close confinement’ of substrate in the cis cavity affects rates of folding by tail-multiplied GroEL. *Proc. Natl. Acad. Sci. USA* 104, 5342–5347.
- Fenton, W.A. and Horwich, A.L. (2003). Chaperonin-mediated protein folding: fate of substrate polypeptide. *Quart. Rev. Biophys.* 36, 229–256.
- Ferreon, A.C.M., Ferreon, J.C., Wright, P.E., and Deniz, A.A. (2013). Modulation of allostery by protein intrinsic disorder. *Nature* 498, 390–394.
- Fierz, B., Satzger, H., Root, C., Gilch, P., Zinth, W., and Kiefhaber, T. (2007). Loop formation in unfolded polypeptide chains on the picoseconds to microseconds time scale. *Proc. Natl. Acad. Sci. USA* 104, 2163–2168.
- Flory, P. (1949). The configuration of real polymer chains. *J. Chem. Phys.* 17, 303.
- Grosberg, A. and Kuznetsov, D. (1992). Quantitative theory of the globule-to-coil transition. 4. Comparison of theoretical results with experimental data. *Macromolecules* 25, 1996–2003.
- Ha, B.-Y. and Thirumalai, D. (1992). Conformations of a polyelectrolyte chain. *Phys. Rev. A* 46, R3012–R3015.
- Haran, G. (2012). How, when and why proteins collapse: the relation to folding. *Curr. Opin. Struct. Biol.* 22, 14–20.
- Hartl, F.U. and Hayer-Hartl, M. (2002). Protein folding – molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* 295, 1852–1858.

- Hartl, F.U., Bracher, A., and Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. *Nature* 475, 324–332.
- Hayer-Hartl, M. and Minton, A.P. (2006). A simple semiempirical model for the effect of molecular confinement upon the rate of protein folding. *Biochemistry* 45, 13356–13360.
- Higgs, P.G. and Joanny, J.-F. (1991). Theory of polyampholyte solutions. *J. Chem. Phys.* 94, 1543–1554.
- Hillger, F., Hänni, D., Nettels, D., Geister, S., Grandin, M., Textor, M., and Schuler, B. (2008). Probing protein-chaperone interactions with single-molecule fluorescence spectroscopy. *Angew. Chem. Int. Ed. Engl.* 47, 6184–6188.
- Hoffmann, A., Kane, A., Nettels, D., Hertzog, D., Baumgärtel, P., Lengefeld, J., Reichardt, G., Horsley, D., Seckler, R., Bakajin, O., et al. (2007). Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proc. Natl. Acad. Sci. USA* 104, 105–110.
- Hofmann, H., Golbik, R., Ott, M., Hübner, C., and Ulbrich-Hofmann, R. (2008). Coulomb forces control the density of the collapsed unfolded state of barstar. *J. Mol. Biol.* 376, 597–605.
- Hofmann, H., Hillger, F., Pfeil, S., Hoffmann, A., Streich, D., Haenni, D., Nettels, D., Lipman, E., and Schuler, B. (2010). Single-molecule spectroscopy of protein folding in a chaperonin cage. *Proc. Natl. Acad. Sci. USA* 107, 11793–11798.
- Hofmann, H., Soranno, A., Borgia, A., Gast, K., Nettels, D., and Schuler, B. (2012). Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proc. Natl. Acad. Sci. USA* 109, 16155–16160.
- Houry, W.A., Frishman, D., Eckerskorn, C., Lottspeich, F., and Hartl, F.U. (1999). Identification of in vivo substrates of the chaperonin GroEL. *Nature* 402, 147–154.
- Jewett, A.I., Baumketner, A., and Shea, J.E. (2004). Accelerated folding in the weak hydrophobic environment of a chaperonin cavity: creation of an alternate fast folding pathway. *Proc. Natl. Acad. Sci. USA* 101, 13192–13197.
- Kerner, M.J., Naylor, D.J., Ishihama, Y., Maier, T., Chang, H.-C., Stines, A.P., Georgopoulos, C., Frishman, D., Hayer-Hartl, M., Mann, M., et al. (2005). Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell* 122, 209–220.
- Klimov, D.K., Newfield, D., and Thirumalai, D. (2002). Simulations of beta-hairpin folding confined to spherical pores using distributed computing. *Proc. Natl. Acad. Sci. USA* 99, 8019–8024.
- Kohn, J., Millett, I., Jacob, J., Zagrovic, B., Dillon, T., Cingel, N., Dothager, R., Seifert, S., Thiagarajan, P., Sosnick, T., et al. (2004). Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. USA* 101, 12491–12496.
- Kuzmenkina, E., Heyes, C., and Nienhaus, G. (2005). Single-molecule Förster resonance energy transfer study of protein dynamics under denaturing conditions. *Proc. Natl. Acad. Sci. USA* 102, 15471–15476.
- Lindorff-Larsen, K., Piana, S., Dror, R.O., and Shaw, D.E. (2011). How fast-folding proteins fold. *Science* 334, 517–520.
- Magg, C., Kubelka, J., Holtermann, G., Haas, E., and Schmid, F.X. (2006). Specificity of the initial collapse in the folding of the cold shock protein. *J. Mol. Biol.* 360, 1067–1080.
- Merchant, K., Best, R., Louis, J., Gopich, I., and Eaton, W. (2007). Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proc. Natl. Acad. Sci. USA* 104, 1528–1533.
- Mittal, J. and Best, R.B. (2008). Thermodynamics and kinetics of protein folding under confinement. *Proc. Natl. Acad. Sci. USA* 105, 20233–20238.
- Möglich, A., Joder, K., and Kiefhaber, T. (2006). End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation. *Proc. Natl. Acad. Sci. USA* 103, 12394–12399.
- Müller-Späth, S., Soranno, A., Hirschefeld, V., Hofmann, H., Rügger, S., Reymond, L., Nettels, D., and Schuler, B. (2010). Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. USA* 107, 14609–14614.
- Nettels, D., Gopich, I., Hoffmann, A., and Schuler, B. (2007). Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc. Natl. Acad. Sci. USA* 104, 2655–2660.
- Neuweiler, H., Schulz, A., Böhmer, M., Enderlein, J., and Sauer, M. (2003). Measurement of submicrosecond intramolecular contact formation in peptides at the single-molecule level. *J. Am. Chem. Soc.* 125, 5324–5330.
- Neuweiler, H., Johnson, C., and Fersht, A. (2009). Direct observation of ultrafast folding and denatured state dynamics in single protein molecules. *Proc. Natl. Acad. Sci. USA* 106, 18569–18574.
- Nozaki, Y. and Tanford, C. (1970). The solubility of amino acids, diglycine, and triglycine in aqueous guanidine hydrochloride solutions. *J. Biol. Chem.* 245, 1648–1652.
- Sanchez, I. (1979). [Phase transition behavior of the isolated polymer chain](#). *Macromolecules* 12, 980–988.
- Schuler, B. (2007). Application of single molecule Förster resonance energy transfer to protein folding. *Meth. Mol. Biol.* 350, 115–138.
- Schuler, B. and Hofmann, H. (2013). Single-molecule spectroscopy of protein folding dynamics-expanding scope and timescales. *Curr. Opin. Struct. Biol.* 23, 1–12.
- Schuler, B., Lipman, E., and Eaton, W. (2002). Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* 419, 743–747.
- Shakhnovich, E.I. and Gutin, A.M. (1989). Formation of unique structure in polypeptide chains: theoretical investigation with the aid of a replica approach. *Biophys. Chem.* 34, 187–199.
- Shakhnovich, E.I. and Gutin, A.M. (1990). Implications of thermodynamics of protein folding for evolution of primary sequences. *Nature* 346, 773–775.
- Sherman, E. and Haran, G. (2006). Coil-globule transition in the denatured state of a small protein. *Proc. Natl. Acad. Sci. USA* 103, 11539–11543.
- Sinha, K.K. and Udgaonkar, J.B. (2005). Dependence of the size of the initially collapsed form during the refolding of barstar on denaturant concentration: evidence for a continuous transition. *J. Mol. Biol.* 353, 704–718.
- Sirur, A. and Best, R.B. (2013). Effects of interactions with the GroEL cavity on protein folding rates. *Biophys. J.* 104, 1098–1106.
- Soranno, A., Buchli, B., Nettels, D., Cheng, R., Müller-Späth, S., Pfeil, S., Hoffmann, A., Lipman, E., Makarov, D., and Schuler, B. (2012). Quantifying internal friction in unfolded and intrinsically disordered proteins with single molecule

- spectroscopy. *Proc. Natl. Acad. Sci. USA* **109**, 17800–17806.
- Stan, G., Brooks, B.R., Lorimer, G.H., and Thirumalai, D. (2006). Residues in substrate proteins that interact with GroEL in the capture process are buried in the native state. *Proc. Natl. Acad. Sci. USA* **103**, 4433–4438.
- Stigter, D. and Dill, K.A. (1990). Charge effects on folded and unfolded proteins. *Biochemistry* **29**, 1262–1271.
- Takagi, F., Koga, N., and Takada, S. (2003). How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: molecular simulations. *Proc. Natl. Acad. Sci. USA* **100**, 11367–11372.
- Tang, Y.C., Chang, H.C., Roeben, A., Wischnewski, D., Wischnewski, N., Kerner, M.J., Hartl, F.U., and Hayer-Hartl, M. (2006). Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein. *Cell* **125**, 903–914.
- Thirumalai, D. and Ha, B.-Y. (1998). *Statistical Mechanics of Semiflexible Chains* (San Diego: Academic Press).
- Thirumalai, D. and Lorimer, G.H. (2001). Chaperonin-mediated protein folding. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 245–269.
- Thirumalai, D., O'Brien, E., Morrison, G., and Hyeon, C. (2010). Theoretical perspectives on protein folding. *Annu. Rev. Biophys.* **39**, 159–183.
- Uversky, V.N. (2002). Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* **11**, 739–756.
- Uversky, V., Gillespie, J., and Fink, A. (2000). Why are 'natively unfolded' proteins unstructured under physiologic conditions? *Proteins* **41**, 415–427.
- Zhou, H.X. and Dill, K.A. (2001). Stabilization of proteins in confined spaces. *Biochemistry* **40**, 11289–11293.
- Ziv, G. and Haran, G. (2009). Protein folding, protein collapse, and Tanford's transfer model: lessons from single-molecule FRET. *J. Am. Chem. Soc.* **131**, 2942–2947.



Hagen Hofmann studied Biochemistry at the Martin-Luther University Halle-Wittenberg. During his time as a PhD student in the group of Professor Renate Ulbrich-Hofmann, he used single-molecule fluorescence spectroscopy and NMR-spectroscopy to understand the collapse and folding of the ribonuclease inhibitor barstar. After receiving his PhD, Hagen joined the group of Professor Benjamin Schuler at the University of Zurich where he focuses on the polymeric properties of unfolded and intrinsically disordered proteins and chaperone-mediated protein folding reactions using single-molecule fluorescence spectroscopy.